Bibliometric analysis: Learning using generative AI

Putri Khoirin Nashiroh*, Ranu Iskandar

Universitas Negeri Semarang, Indonesia

Abstract: This research aims to 1) analyze the number of articles per year and their types, 2) analyze the five articles with the most citations, 3) analyze keywords in the data collected, and 4) analyze the relationship between authors. Bibliometric analysis was used to help researchers study bibliographic content and citation analysis of each article taken from Harzing’s Publish or Perish (PoP) 8 database. The data was taken from Scopus January 2018-2024 entitled, “learning using generative AI” in English. The maximum number of articles accessed is 500 articles. The collected articles are stored in CSV and RIS format. Articles were filtered according to analysis needs using MS Excel and VOSViewer. Twenty-nine documents from the Scopus database have been used in this research. The results show that 1) in 2022, there was one manuscript; in 2023, there were 27 manuscripts; and in 2024, there was one manuscript published. There are ten manuscripts published in conference proceedings, 14 articles in journals, and five manuscripts in edited books; 2) Suh, Popenici, Crosthwaite, Kim, and Chan wrote the 5 most cited articles; 3) Future research topics include GPT, student agency, learning approaches, higher education, state-of-the-art, reinforcement learning, opportunities, and challenges. Potential future research (novelty) could use keywords that are not yet widely used, including pedagogy, simulations, Socratic tutors, teaching methods, neuro-inclusive learning, and metacognition; 4) a network of writers outside the group consisting of Han, Ariel; Leu, U; Leu, Eunso; Lim, Cheoil; Kim, Hyeoncheol; Lee, Jeongji; Kim, Jiwon is the cluster with the highest relationship between nodes.

Keywords: Bibliometric, generative AI, learning

*Corresponding author: putrikhoirin@mail.unnes.ac.id
INTRODUCTION

The Industrial revolution (IR) progresses faster in each era, starting from 0.0 to 5.0. Before 1760, people thought traditionally of agricultural triggers for survival. In 1760-1830, society entered IR 1.0, where steam engines and water power mechanism systems emerged, resulting in the emergence of textile, iron, steel, and transportation industries. Starting in the 1870s, society entered IR 2.0, where electrical technologies emerged and the mechanization of iron and steel mass production systems, car industries, and the discovery of oil refineries to replace coal. In the 1970s, IR 3.0 began to enter, where computers, cellphones, and the internet appeared and entered industries. From 2010 until now, we entered IR 4.0, where robot technologies, chatbot artificial intelligence (AI), and the Internet of Things (IoT) emerged (Maghfoiroh & Iskandar, 2023). We are currently transitioning towards IR 5.0, where AI and robots are here to work together with humans. It encourages efficiency and productivity by utilizing technology utilized by human intelligence (Ikhsan, 2023; Restendy et al., 2021; Siagian, 2023; The Editors of Encyclopaedia Britannica, 2023).

The increasingly rapid shift in the industrial revolution means that all sectors, including education, must adapt. If RI 4.0, education 4.0 emerges as the education sector's answer to adapting, then when RI 5.0, education 5.0 will appear (Iskandar, 2024). Education 5.0 is the integration of AI and IoT in the education sector. Artificial intelligence (AI) is a computer program that works like human intelligence to carry out tasks and solve problems (Korteling et al., 2021). AI was introduced by John McCarthy at a summer conference at Dartmouth College in 1956. AI branches include machine learning, deep learning, natural language processing, and computer vision. Generative AI is machine learning capable of producing new content such as text written in natural language, images (including photos, digital paintings, and cartoons), videos, music, and software code in response to commands written in natural-language conversational interfaces (Google, 2024; UNESCO, 2023).

Text generative AI uses a type of ANN known as a general-purpose transformer and a large language model. Text GenAI is often referred to as large language modeling or LLM. The type of LLM used by Text GenAI is known as a generative pre-trained transformer (UNESCO, 2023). An example is chatGPT (chat.openai.com). ChatGPT can answer all questions in the form of text prompts, carry out conversations, and check the similarity of writing (Maghfoiroh & Iskandar, 2023). In its development, various alternatives to chatGPT emerged, including Alpaca, Bard, Chatsonic, Ernie, Hugging Chat, Perplexity, Jasper, Llama, Open Assistant, Tongyi Qianwen, YouChat, ChatPDF, and Elicit. In the development of chatGPT, chatGPT 4 can use image generative AI as a prompt.

Image generative AI and music generative AI use an ANN generative adversarial network (GAN) that can also be combined with variational autoencoders (UNESCO, 2023). Examples of Image-generative AI are Craiyon, DALL E 3, Midjourney, Adobe Firefly, Generative AI by Getty, and Stable Diffusion (Guinness, 2024). Examples of Music generative AI are Aiva, Boomy, Soundraw, Magenta Studio, Synthesizer V, Voicemod, Image to Sound Effect, and Melobytes (Sharma, 2023). Examples of AI generative videos are Elai, GliaCloud, Pictory, and Runway (Alfred, 2023).

Many publications discuss vocational learning with generative AI. Various articles can be accessed via Google Scholar or directly on the Google search engine. Several previous publications have conducted research on learning with generative AI in vocational schools, such as that undertaken by Rahman et al. (2023) stated that using ChatGPT provides...
significant benefits for students and teachers, with ChatGPT’s ability to answer questions, summarize documents, translate text, and understand program code. ChatGPT is effective in meeting the learning needs of students at SMK Negeri 1 Pantai Labu. Arisanti et al. (2024) showed that the use of AI in learning can provide optimal learning experiences, develop creativity and professional skills of educators, design learning strategies that are innovative and responsive to changing educational needs, and increase teaching effectiveness. Zulfikar et al. (2023) showed that Quillbot, a Generative AI, can improve teacher writing skills in the aspects of coherence, grammar, and structure. Although many authors have published publications related to learning with generative AI, until now, no one has carried out mapping using bibliometric analysis of related publications indexed by Scopus with the help of VOSviewer.

The more diverse generative AI becomes, the more researchers are researching generative AI. From the research potential above, the objectives of this research are 1) analyzing the number of articles per year and their types, 2) analyzing the five articles with the most citations, 3) analyzing keywords in the data collected, and 4) analyzing the relationship between authors.

**METHOD**

Analysis is part of the research evaluation method and from the various literature that has been produced, it is possible to carry out bibliometric analysis using its method (Ellegaard & Wallin, 2015). The method used was bibliometric analysis to help researchers in studying bibliographic content and citation analysis of each article taken from the Harzing’s Publish or Perish (PoP) 8 database. The data was taken from Scopus January 2018-2024 with the title, “Learning using generative AI” in English. The maximum number of articles accessed is 500 articles. The collected articles are stored in CSV and RIS format. Articles filtered according to analysis needs. VOSviewer 1.6.18 can visualize predefined article searches; the steps can be summarized as follows: determine search keywords, perform article search, filter and process search result data according to needs, collect and compile statistical data on search results, as well as conduct bibliometric analysis (Iskandar, 2023, Iwan et al., 2023). Meanwhile, according to research Julia et al. (2020), the four stages of the bibliographic study are: (1) search processes; (2) bibliographic filters; (3) complete bibliography; and (4) bibliometric analysis.

Forty-four manuscripts and 72 citations were harvested using PoP 8 until early 2024. The data were saved in CSV format for Microsoft Excel and RIS format for VOSviewer software.

![Fig. 1. Manuscript harvest results using PoP](image-url)
197

CSV file on Ms. Excel and RIS files in Mendeley were selected one by one to see whether the content matches the title learning using generative AI. If something does not match the title, the data is deleted. Of the 44 existing manuscripts, 29 manuscripts were found that matched the title. In Mendeley, the 29 manuscripts were completed with the contents by writing down all the authors and entering author keywords.

Fig. 2. RIS file opened in mendeley

RESULTS AND DISCUSSION

The data in Excel was filtered in tables. The table title "year" was used to group the year the manuscript was published. Type headings were used to group publication types. The results of grouping publication years showed that in 2018-2021, there were no manuscripts published; in 2022, there was one manuscript; in 2023, there were 27 manuscripts published; and in 2024, there was one manuscript published. The results of grouping the forms of published manuscripts show that ten manuscripts were published in conference proceedings, 14 articles in journals, and five manuscripts in edited books.

Fig. 3. The CSV file opened in MS excel

The data in Excel was filtered in tables. The table title "year" was used to group the year the manuscript was published. Type headings were used to group publication types. The results of grouping publication years showed that in 2018-2021, there were no manuscripts published; in 2022, there was one manuscript; in 2023, there were 27 manuscripts published; and in 2024, there was one manuscript published. The results of grouping the forms of published manuscripts show that ten manuscripts were published in conference proceedings, 14 articles in journals, and five manuscripts in edited books.
Table 1. The top 5 manuscripts

<table>
<thead>
<tr>
<th>Cites</th>
<th>Authors</th>
<th>Titles</th>
<th>Years</th>
<th>Sources</th>
<th>Publishers</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Suh, S. &amp; An, P.</td>
<td>Leveraging Generative Conversational AI to Develop a Creative Learning Environment for Computational Thinking</td>
<td>2022</td>
<td>International Conference on Intelligent User Interfaces, Proceedings IUI</td>
<td>ACM</td>
</tr>
<tr>
<td>7</td>
<td>Popenici, S., Rudolph, J., Tan, S., &amp; Tan, S.</td>
<td>A critical perspective on generative AI and learning futures. An interview with Stefan Popenici</td>
<td>2023</td>
<td>Journal of Applied Learning and Teaching</td>
<td>Simon Fraser University</td>
</tr>
<tr>
<td>2</td>
<td>Chan, C.K.Y &amp; Lee, K. K. W.</td>
<td>The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers?</td>
<td>2023</td>
<td>Smart Learning Environments</td>
<td>Springer Nature</td>
</tr>
</tbody>
</table>

From the data above, it can be seen that 80% of the articles cited came from journals, while the rest came from conference proceedings. However, articles published in conference proceedings had the highest number of citations, with eight citations. This article can be accessed at the ACM Digital Library. This is because articles published in conference proceedings are open access, whereas, in journals, there is usually a fee to access them (AlRyalat et al., 2019; Pranckutė, 2021).

The edited RIS data was then opened using VosViewer. The analysis used in VosViewer was based on keywords and authors. The analysis procedure was to click "create," and then a popup appears to select the type of data. Click "Create a map based on bibliographic data" and click "Next." Then, a pop-up would appear to choose the data source. Clicked "read data from reference management files" and click "Next." Then, a pop-up would appear to select the file. Clicked ris and select the ris file that has been edited using Mendeley, then click "next." Then, a pop-up would appear to choose the type of analysis and counting method. There are two types of analysis methods, namely co-authorship and co-occurrence. Co-authorship is used to analyze authors writing in the same network, whether directly or indirectly. Co-occurrence is used to analyze keywords (University of Illinois Chicago, 2023). Choose one analysis method. When you have finished visualizing the data, choose another analysis method. In the counting method, select full counting and click "next." Then, a pop-up would appear to select the threshold. Wrote the number 1 so that all
authors or keywords appear in the visualization, then click "next." Wrote down the number of all authors or keywords and clicked "finish."

Fig. 4. Network visualization of co-occurrence

Co-occurrence analysis displays network visualization between keywords. The larger the node, the greater the number of keywords that appear (Zakaria et al., 2022). Edge connects between nodes and the strength of the relationship. The closer the distance between one node and another node is, the higher the relationship between the nodes is (Aribowo, 2019). Figure 4 shows a network visualization of co-occurrence. The keyword that appears most often is generative AI. Colored nodes indicate how many clusters there are. There are 13 clusters in the network visualization of co-occurrence. Cluster 1-8 is the dominant cluster because the keywords appear a lot, while cluster 9-1 is the dominant cluster because the keywords rarely appear.

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (red)</td>
<td>Advantages, AI literacy, benefits, challenges, competencies, holistic, opportunities, reinforcement learning, risks, state-of-the-art</td>
</tr>
<tr>
<td>2 (green)</td>
<td>auto-generated course units, bias, diffusion model, education, equality, generative AI models, language learning, LLM, personalized learning</td>
</tr>
<tr>
<td>3 (blue)</td>
<td>Aided, algorithmic bias, artificial intelligence (AI), big tech, education technology, fascism, higher education, superintelligence</td>
</tr>
<tr>
<td>4 (yellow)</td>
<td>Editorial, genai, learning analytics, practice, research, user-centered research</td>
</tr>
<tr>
<td>5 (purple)</td>
<td>AI in education, ai-generated learning content, impact</td>
</tr>
<tr>
<td>6 (arctic)</td>
<td>Autonomous learning, Flipped classroom, GPT, learning approaches, student agency.</td>
</tr>
<tr>
<td>7 (orange)</td>
<td>Learning, new literacies, post-primary students, python programming</td>
</tr>
<tr>
<td>8 (grape)</td>
<td>Chatbot, gpt-4, quizizz ai, self assessment</td>
</tr>
</tbody>
</table>
Figure 5 displays an overlay visualization of co-occurrence. Visualization of co-occurrence aims to intuitively understand the research development trends and predict future orientations and hotspots (Martínez-Heredia et al., 2022). This figure displays possible future research trends. The bluer the node, the earlier the year of publication; conversely, the more recent the publication, the yellower the node. Possible future research topics include GPT, student agency, learning approaches, higher education, state-of-the-art, reinforcement learning, opportunities, and challenges.

Figure 6 displays a density visualization of co-occurrence. This image is a topic that still needs to be used regarding generative AI. Hotspots on the map that are increasingly
yellow indicate that more keywords are being used, while the darker the color, the more research using these keywords is rarely carried out (Soesanto & Handalani, 2023). Potential future research could use keywords that have yet to be widely used, including pedagogy, simulations, Socratic tutors, teaching methods, neuro-inclusive learning, and metacognition.

Co-authorship analysis to see author networks. Co-authorship analysis is useful to see the mapping of research topics through relationships or collaborations between authors (Putri et al., 2023). In this study, 65 authors from 29 manuscripts met the requirements for analysis. Which means there are 2.24 authors per manuscript. Data visualization shows that cluster 1 has a network of writers outside the group consisting of Han, Ariel; Leu, U; Leu, Eunso; Lim, Cheoil; Kim, Hyeoncheol; Lee, Jeongji; Kim, Jiwon is the cluster with the highest relationship between nodes. Other authors collaborate with teammates or even research individually.

CONCLUSION

The development of learning using generative AI in the last five years has been analyzed using bibliometric methods. In 2022, there was one manuscript; in 2023, there were 27 manuscripts; and in 2024, there was one manuscript published. There are ten manuscripts published in conference proceedings, 14 articles in journals, and five manuscripts in edited books. Suh, Popenici, Crosthwaite, Kim, and Chan wrote the 5 most cited articles. Future research topics include GPT, student agency, learning approaches, higher education, state-of-the-art, reinforcement learning, opportunities, and challenges. Potential future research (novelty) could use keywords that are not yet widely used, including pedagogy, simulations, socratic tutors, teaching methods, neuro-inclusive learning, and metacognition.
Leu, U; Leu, Eunso; Lim, Cheoil; Kim, Hyeoncheol; Lee, Jeongji; Kim, Jiwon is the cluster with the highest relationship between nodes.

REFERENCES


Julia, J., Afrianti, N., Soomro, K. A., Supriyadi, T., Dolifah, D., Isrokutun, I., Erhamwilda, E., &


Suh, S., & An, P. (2022). Leveraging Generative Conversational AI to Develop a Creative Learning Environment for Computational Thinking. In International Conference on
Intelligent User Interfaces, Proceedings IUI (pp. 73–76). https://doi.org/10.1145/3490100.3516473


